Journal of Korean Society for Atmospheric Environment Vol. 37, No. 6, December 2021, pp. 862-870 https://doi.org/10.5572/KOSAE.2021.37.6.862 p-ISSN 1598-7132, e-ISSN 2383-5346

논 문



인공신경망 모델과 배경대기 측정자료를 활용한 서울시 PM_{2.5} 농도 단기예측 및 입력변수의 기여도 분석

Calculation of PM_{2.5} in Seoul 12-hours in Advance Using Simple Artificial Neural Network with Measurements of Background Sites, and Analysis of Contribution of Input Variables

길준수, 이미혜*

고려대학교 이과대학 지구환경과학과

Junsu Gil, Meehye Lee*

Department of Earth and Environmental Sciences, Korea University, Seoul, Republic of Korea

접수일 2021년 8월 6일 수정일 2021년 9월 20일 채택일 2021년 11월 23일

Received 6 August 2021 Revised 20 September 2021 Accepted 23 November 2021

*Corresponding author Tel : +82-(0)2-3290-3645 E-mail : meehye@korea.ac.kr

Abstract Recently, Artificial Neural Network (ANN) models have been successfully applied to predict $PM_{2.5}$ mass concentration. However, the complex nature of ANNs hinders understanding of the actual relationship between input variables and output $PM_{2.5}$. In this study, a simple ANN model was constructed to predict the $PM_{2.5}$ mass of Seoul 12 hours in advance using nine atmospheric variables routinely measured in Seoul and three Background sites. The contribution of the input variables from the four sites to the predicted $PM_{2.5}$ mass was then estimated using the Connection Weight Method (CWM) and the Garson's Algorithm (GA). The second rank of Baengnyeong Island $PM_{2.5}$ after Seoul suggests the impact of transport, and the least contribution of reactive gases of Seoul including O_3 , NO_2 , SO_2 , and CO, indicates the relatively insignificant contribution of *in situ* formation to $PM_{2.5}$. The ranking of meteorological variables including temperature, relative humidity, and wind direction and speed highlights the importance of synoptic meteorological conditions in determining $PM_{2.5}$ levels in Seoul. It also reveals the role of stagnation in increasing $PM_{2.5}$ mass.

Key words: Artificial neural network, PM_{2.5} prediction, Input variable feature importance

1. 서 론

초미세먼지 (PM_{2.5})는 2015년 전국적으로 공식적 인 관측이 시작된 이후, 연평균 농도의 변화는 크지 않으나 고농도 사례 빈도가 2019년까지 지속적으로 증가하였다. 이에 2018년 3월 대기환경기준이 강화 되었고, 비상저감조치와 계절관리제 등을 시행하며 특히 고농도가 주로 발생하는 12~3월에 농도 저감과 예보 정확도 향상을 위한 집중적인 노력을 기울여 왔 다. 국립환경과학원에서는 하루 4회 PM_{2.5} 농도를 4 등급으로 예보하는데 정확도 향상을 위해 화학수송 모델의 개선과 더불어 최근에는 인공신경망 모델을 보조적으로 활용하고 있다(Ho *et al.*, 2021; Chang *et al.*, 2016). 기상 변화가 역동적이고 지형이 복잡한 우 리나라 여건상 통계적 방법에 의한 PM_{2.5} 예측이 비 교적 높은 정확도를 보여준다(Cho *et al.*, 2019a; Cho *et al.*, 2019b). 만일 인공신경망 모델에서 입력자료와 PM_{2.5} 간의 상관관계를 파악할 수 있다면 기존의 연 역적 방법에 기반한 모델에서 모사하지 못하는 물리 화학적 현상들을 이해하는 데 도움이 되어 예보 모델 개선에 기여할 수 있을 것이다 (Reichstein *et al.*, 2019). 그러나 인공신경망 모델은 내부 구조의 복잡 성으로 인하여 물리적으로 유의미한 인과관계를 파 악하기 매우 어렵다 (Adadi and Berrada, 2018).

따라서 인공신경망 모델 자체를 이해하기 위한 여 러가지 기법이 발전되어 왔으며, 그중 "가중치 연결 법" (Connection Weight Method; CWM)과 "가르손 알고리즘" (Garson's Algorithm; GA) 방법이 주로 사 용되고 있다 (Olden *et al.*, 2004). CWM은 모델에 사 용된 입력 자료가 결과값에 주는 영향을 상대적인 순 위로 산출할 수 있으며, GA는 CWM에 비해 순위의 정확도는 낮지만 영향 정도를 정량적으로 얻을 수 있 는 장점이 있다. 이러한 기법을 이용하면, 여러 종류 의 대기 관측 자료를 함께 사용하여 PM_{2.5} 질량 농도 를 예측하는 모델을 구축할 때, 각 인자의 영향 파악 이 가능하다. 더불어 복잡하게 연결된 인자 간의 관 계에 대한 이해를 향상시켜, 궁극적으로 예보 성능을 향상시키는 데 기여할 수 있다.

본 연구에서는 국가 배경지역, 즉 서해의 백령도와 제주도, 그리고 동해의 울릉도 세 곳의 PM_{2.5} 질량농도 와 반응성기체 그리고 기상인자 측정 자료와 서울의 자료를 사용하여 12시간 후 서울시 PM_{2.5} 농도를 모사 하였다. 이후 각 지역별로 PM_{2.5}, 반응성기체와 기상인 자의 기여도를 분석하였다. 이를 위해 먼저 12시간 후 서울시의 PM_{2.5} 질량농도 산출을 위한 인공신경망 모 델을 구축하였으며, 이를 기반으로 CWM과 GA 방법 을 활용하여 각 지역의 입력 인자별 기여율을 정성, 정 량적으로 산정하여 비교하고 그 의미를 분석하였다.

2. 연구 방법

2.1 입력자료 구축

환경부에서는 2002년 이래 전국에 584개소의 대기 환경측정망을 설치하여 운영하고 있다. 각 측정소에 서는 반응성 기체 (O₃, NO₂, CO, SO₂), 미세먼지 (PM₁₀), 초미세먼지 (PM_{2.5})가 실시간으로 측정되며 1 시간 평균 자료가 웹사이트를 통해 공개된다(http:// www.airkorea.or.kr/web).

본 연구에서는 서울시의 5개권역 중 도심권에 속 하는 중구의 대기환경측정소(S; 37.56°N, 126.98°E) 와 국가배경대기관측소인 백령도(BI; 37.95°N, 124.63°E), 제주도(Ji; 33.29°N, 126.16°E), 울릉도(UI; 37.51°N, 130.82°E)에서 2015년에서 2019년까지 5년 동안 측정된 1시간 평균자료를 사용하였다(그림 1). 백령도, 제주도, 울릉도는 서울로부터 약 210~460 km 거리에 있으며, 계절에 따라 서울의 풍상과 풍하 지역에 위치한다(그림 2)(Lim *et al.*, 2016; Lim *et al.*, 2013; Moon *et al.*, 2005; Park *et al.*, 1994). 기상자료로 는 같은 기간 동일 지역의 종관기상 측정소에서 1시 간 간격으로 산출되는 온도(T), 상대습도(RH), 풍속 (WS), 풍향(WD) 자료를 활용하였다(https://data. kma.go.kr/cmmn/main.do).

PM_{2.5} 농도는 겨울과 여름의 계절에 따른 차이가 분명하게 나타나므로 이를 고려하여 여름(6~8월)과 겨울(12~2월)로 구분하여 모델을 구축하고 분석하 였다. 풍향은 각도(°)의 cosine 값을 사용하였으며, 광 화학 반응을 고려하여 시간을 밤(18:00~06:00)과 낮 (06:00~18:00)으로 구분하여(밤: 0, 낮: 1) 입력자료에 추가하였다. 측정값은 단위가 다르고 값의 범위가 다 르므로 최소-최대값에 대한 비율(min-max scaling method)로 정규화한 후 모델에 입력자료로 사용하였 다(식 1).

$$x_{nor} = \frac{x - minimum(X)}{maximum(X) - minimum(X)}$$
(1)

입력자료는 무결성을 위하여, 측정값이 하나라도 누락되지 않은 시간의 자료로만 구축되었다. 입력자 료의 확보율은 서울이 가장 높았고 제주가 가장 낮았 다(표 1).

2.2 인공신경망 모델 구성

일반적으로 입력층(input layer; x) 내 i개의 입력 노 드(input feature)와 1개의 숨김층(hidden layer) 내 j개의 노드를 가진 인공신경망 모델에서(그림 3),

J. Korean Soc. Atmos. Environ., Vol. 37, No. 6, December 2021, pp. 862-870

첫 번째 숨김층의 1번째 노드(h_1^1)는 아래의 식으로 표현할 수 있다(식 2).

$$h_{1}^{1} = \phi \left(x_{1} w_{1,1}^{x,h^{1}} + x_{2} w_{2,1}^{x,h^{1}} + \dots + x_{i} w_{i,1}^{x,h^{1}} + b_{1}^{x} \right)$$

$$= \phi \left(\sum_{n=1}^{n} x_n w_{n,1}^{x,n} + b_1^x \right)$$
 (2)

위 식에서 $W_{i,1}^{x,h^1}$ 는 입력층 (input layer; χ)의 i번째



Fig. 1. The map shows four observation sites including Seoul, Baegnyeong Island, Jeju Island, and Ulleung Island.

노드 (x_i) 와 첫 번째 숨김층의 첫 번째 노드 (h_1^1) 간 가중치, Φ 는 활성함수, b_1^x 는 편향항(bias term)을 나타낸다. 이 편향항을 $x_0=1$, $w_{0,1}^{x,h^1}=b_1^x$ 이라 간 주하면 아래 식으로 표현된다(식 3).

$$b_1^x = 1 \times b_1^x = x_0 w_{0,1}^{x,h^1} \tag{3}$$

식 3을 식 2에 적용하였을 때, 첫 번째 숨김층의 1 번째 노드를 의미하는 식 2의 *h*₁¹은 최종적으로 아 래와 같이 표현된다(식 4).

$$h_1^1 = \Phi\left(\sum_{n=0}^i x_n w_{n,1}^{x,h^1}\right)$$
(4)

따라서 모든 노드를 고려하여 일반항으로 표현하

 Table 1. Acquisition rate (%) of input variables at four observation sites.

	Seoul	Beagnyeong	Ulleung	Jeju
PM _{2.5}	98.8	92.6	80.5	75.8
O ₃	99.1	96.8	90.8	95.4
NO ₂	96.7	97.3	89.6	92.1
CO	95.5	97.3	90.5	77.4
SO ₂	98.7	96.8	85.7	85.6
Т	100.0	100.0	100.0	100.0
RH	99.9	100.0	99.9	99.9
WS	99.9	100.0	100.0	99.6
WD	99.8	99.8	99.8	99.1



Fig. 2. 12-hours backward trajectories of air masses arriving at Seoul in (a) winter and (b) summer from 2015 to 2019.



Fig. 3. The structure of Artificial Neural Network (ANN) model used in this study. Input variables were divided into four groups including PM2.5, precursor gases, meteorological parameters, and photochemical surrogate.

면, 첫 번째 숨김층의 j번째 노드(h_i)는(식 5)로, 출 산할 수 있다. 력층 내 유일한 노드인 출력값 (y1)은 (식 6)으로 표 혀된다.

$$h_{j}^{1} = \phi\left(\sum_{n=0}^{i} x_{n} w_{n,j}^{x,h^{1}}\right)$$
(5)

$$y_1 = \phi\left(\sum_{m=0}^{j} h_m^1 w_{m,1}^{h^1, y}\right)$$
(6)

2.3 가중치 연결법(Connection Weight Method; CWM)과 가르손 알고리즘 (Garson Algorithm; GA)

"가중치 연결법" (Connection Weights Method; CWM)은 시뮬레이션 연구 결과를 통해 입력자료의 기여도를 판별하는데 정확도가 가장 높은 것으로 보 고되었다(Olden et al., 2004; Olden and Jackson, 2002). 가중치 연결법을 통한 입력층 내 1번째 노드 (*x*₁)의 출력값(*y*₁)에 대한 기여도는 아래 식 7로 계

$$P_{x_1 \to y_1} = w_{1,1}^{x,h^1} \times w_{1,1}^{h^1,y} + w_{1,2}^{x,h^1} \times w_{2,1}^{h^1,y} + \cdots + w_{1,j}^{x,h^1} \times w_{j,1}^{h^1,y} = \sum_{m=1}^j w_{1,m}^{x,h^1} \times w_{m,1}^{h^1,y}$$
(7)

위 식을 통해 계산한 $P_{x_1 \rightarrow y_1}$ 는 노드 간 연결 가중치 곱의 총합을 의미하며, 출력값 (y1)에 대한 입력값 (X1)의 기여도를 정성적으로 나타낸다. 따라서 각 입력 변수들의 기여도 $(P_{x_i \rightarrow y_i})$ 를 산출하여 비교를 통해 입력 노드의 기여도를 정성적으로 판별한다. "가르손 알고리즘" (Garson Algorithm; GA)에서는 노드 간 연결 가중치 절대값의 상대적 비율을 계산 하여, 입력값의 기여도를 정량적으로 계산하며 $P_{x_1 \to y_1}$ 은 식 8을 통해 얻을 수 있다(Goh, 1995; Garson, 1991).

J. Korean Soc. Atmos. Environ., Vol. 37, No. 6, December 2021, pp. 862-870



Fig. 4. The index of agreement (IOA) against node number.

$$P_{x_1 \to y_1} = \sum_{m=1}^{j} \frac{\left| w_{1,m}^{x,h^1} \times w_{m,1}^{h^1,y} \right|}{\sum_{n=1}^{i} \left| w_{n,m}^{x,h^1} \times w_{m,1}^{h^1,y} \right|}$$
(8)

2.4 인공신경망 모델의 성능 평가

인공신경망 모델은 파이썬 기반 tensorflow-gpu 라 이브러리를 이용해 구축하였다(*Python* v3.5, *Tensorflow* v2.3.0, *CUDNN* v7.6.5, *Cuda* v10.1). 활성함수로 는 ELU를, 손실함수로는 MSE를 사용하였다. Batch 와 Epoch은 각각 32와 50으로 설정하였다. 모델의 성 능 평가는 Index Of Agreement (IOA)와 and Mean Absolute Error (MAE)를 사용하였다(식 9, 10).

$$IOA = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (|P_i - \bar{O}| + |O_i - \bar{O}|)^2}$$
(9)

$$MAE = \frac{\sum_{i=1}^{n} |O_i - P_i|}{n}$$
(10)

이때 O_i 는 측정값, P_i 는 예측값이며, \overline{O} 는 측정값의 평균이다.

일반적으로 인공신경망 모델의 성능 검증에는 알 려지지 않은 자료 (unknown data)를 사용하며, 대표 적인 방법으로 k-교차검증(k-fold cross validation)이 있다(Bengio and Grandvalet, 2003). 이는 모델의 과 적합을 피하고 범용적인 예측 능력을 확보하여, 일반 적인 인공신경망 모델의 요구사항인 미지의 자료에

Table 2. The number of nodes for maximum IOA (IOA_{max}) and minimum MAE (MAE_{min}) during the summer and winter.

	Summer	Winter
Node _{IOAmax}	12	16
IOA _{max}	0.81	0.86
MAE (μg m ⁻³)	6.44	7.15

대한 정확한 예측 능력을 높이기 위함이다. 하지만 모델의 내부 구조를 분석하기 위해 입력 노드의 기여 도를 계산하는 경우, 과적합 모델이 자료의 구조를 더 상세하게 반영할 수 있다. 따라서 본 연구에서는 전체 자료를 훈련에 사용하였고, 알려진 자료(known data)를 validation에 재사용하여 해당 자료에 과적합 된 인공신경망 모델을 구축하였다.

모델의 성능평가는 은닉층의 노드 개수를 늘려가 며 수행하였는데, 계절별로 각각 특정 노드 개수에서 가장 높은 성능을 보였다(그림 4, 표 2). 여름에는 노 드가 12개일 때 IOA = 0.81로, 그리고 겨울에는 노드 가 16개일 때 IOA = 0.86으로 최대치를 보였다. 이때 MAE는 각각 6.44 µg m⁻³과 7.15 µg m⁻³이었다(그림 5). 전체적인 PM_{2.5} 농도의 추세 예측 성능은 높았으 나, 배경대기 자료를 사용한 등의 이유로 피크 농도 모사 시 모델의 예측 결과가 실제 측정 값보다 다소 낮게 나타났다.



Fig. 5. Hourly observation and model prediction result of PM_{2.5} in (a) summer and (b) in winter.

3. 결과 및 고찰

3.1 기여도 산출과 평가

모델의 출력 값인 계절별 12시간 후 서울의 PM_{2.5} 질량 농도에 대한 입력 노드의 기여도를 전체 입력변 수를 3개의 그룹으로 구분하여 CWM와 GA 방법으 로 각각 계산하였다. 각 지역에서 PM_{2.5}와의 직접적 인 기여도 (Factor 1), 전구 기체와의 이차생성을 통한 기여도 (Factor 2), 그리고 기상요소를 통한 종관 또는 특정 기상의 기여도 (Factor 3)를 분석하였다.

여름과 겨울에 대해 각 사이트의 3가지 factor의 기 여도를 CWM 방법으로 계산한 후 이를 인자별 순위 로 나타내어 비교하였다(표 3). 3개 Factor에 대해 4 개 지역의 순위는 각기 달랐지만 계절별로는 큰 차이 를 보이지 않았으며 상위와 하위 순위로 구분하여 비 교하면 두 계절의 차이는 없었다. 특히 Factor 1은 4 개 지역의 순위가 두 계절 모두 같았는데, 서울의

Table 3. The contribution rank of feature importance (three factors) for summer and winter at four stations calculated using CWM method.

867

Season		Factor 1	Factor 2	Factor 3
	S	1	4	3
	BI	2	3	1
Summer	UI	4	1	2
	JI	3	2	4
	S	1	3	3
	BI	2	4	2
Winter	UI	4	2	1
	JI	3	1	4

Factor 1 기여도가 가장 크고 그 다음 백령도의 Factor 1 기여도가 높았다. 이와 반대로 Factor 2는 울릉도와 제주도의 기여도가 높았다. Factor 3은 백령도와 울릉 도의 기여도가 높았으며, 서울과 제주의 기여도는 낮 았다.

Factor 1은 12시간 전 PM2.5 농도와의 상관성이므

868

로 서울 자체의 기여가 가장 크고 울릉도의 영향이 미미한 것은 매우 당연하다. 하지만 겨울은 물론이고 여름에도 백령도의 기여도가 두 번째로 큰 것은 국외 로부터의 유입의 영향을 시사하는 것으로 볼 수 있 다. Factor 2는 반응성 기체(O₃, NO₂, CO와 SO₂)인데, 이 중 SO₂는 농도 변화가 크지 않고 CO는 대기 중 체 류 시간이 길어 지역별 차이가 크지 않아 주로 NO2 와 O₃의 영향으로 판단된다. Factor 2는 Factor 1과 반 대의 경향을 보였으며, 특히 여름에는 서울의 기여도 가 가장 낮아 정반대의 순위를 보였다. 겨울에는 백 령도의 기여도가 가장 작고 제주의 기여도가 가장 컸 다. 이러한 계절적 차이는 PM2.5 생성 시 NO2와 O3의 중요성을 반영하며 산화 기작을 통한 NO2와 O3의 영 향은 선행 연구에서도 제시된 바 있다(Wang et al., 2019). 따라서 Factor 2는 PM25에 대한 이차생성의 영향을 나타내는 것으로 간주할 수 있다. 특히 서울 은 Factor 1 기여도가 가장 높으면서 Factor 2의 기여 도가 작았으므로 이는 서울의 자체적인 이차생성의 영향이 타지역에 비해 상대적으로 크지 않음을 암시 한다. 기상변수인 Factor 3은 상위와 하위로 나누면, 두 계절 모두 백령도와 울릉도의 기여도가 크며, 제 주도의 Factor 3 영향이 가장 낮았다. 이러한 Factor 3 영향의 차이는 울릉도 및 백령도와 제주도의 지역 간 종관기상 차이로부터 기인된다(Hwang et al., 2020). 지역 간 종관기상의 차이로 인한 제주도의 낮은 Factor 3 기여도는 제주도의 PM2.5 고농도 발생 특성이 서울과 다름을 보인 사전 연구에서 잘 드러난다(Kim et al., 2020; Kim, 2006). 따라서 Factor 3의 지역별 기 여도 결과는 백령도-서울-울릉도의 전체적인 종관기 상이 서울의 PM2.5 농도에 미치는 영향이 크다는 것 을 의미한다.

지역별 기여율을 GA 방법으로 산출하면, Factor 1 은 서울이 5.7~6.4%로 두 계절 모두 가장 높았으며 여름에는 제주도 (3.7%), 겨울에는 백령도 (4.0%)가 두 번째로 높았다(표 4). Factor 2는 여름에는 울릉도 (13.3%), 제주도 (11.5%) 순으로, 겨울에는 제주도 (13.2%), 백령도 (12.9%) 순으로 기여율이 높았다. 마 **Table 4.** Contribution ratios of the three factors at four sites calculated by GA method.

Season		Factor 1 (%)	Factor 2 (%)	Factor 3 (%)	Total (%)*
	S	5.7	9.1	9.5	24.3
Cumanaan	BI	3.0	8.5	9.4	21.0
Summer	UI	3.1	13.3	11.5	27.8
	١٢	3.7	11.5	8.0	23.2
	S	6.4	9.8	9.3	25.5
\\ <i>\</i>	BI	4.0	12.9	8.9	25.8
winter	UI	3.0	9.6	8.1	20.7
	١٢	4.0	13.2	8.3	25.5

*Sum of the contribution ratio of Factor $1 \sim$ Factor 3

지막 Factor 3의 기여도는 여름에는 백령도(11.5%) 가, 겨울에는 서울(9.3%)이 가장 높았다. GA 방법은 기여도를 정량적으로 산출하므로 세 인자의 종합 기 여도 산출이 가능한데, 여름에는 울릉(27.8%), 서울 (24.3%), 제주(23.2%), 백령(21.0%), 겨울에는 백령 (25.8%), 제주(25.5%), 서울(25.5%), 울릉(20.7%) 순 이었다. 계절별 GA의 총 합은 여름 96.3%, 겨울 97.5%이며, 약 2.5~3.7%는 해석이 불가능한 불확실 도에 해당한다.

GA 방법의 결과로 산출된 Factor 1과 Factor 2의 1 순위는 CWM 결과와 일치하였다. 하지만 Factor 3에 대한 기여도가 특히 겨울에 큰 차이를 보였는데, 이 러한 차이는 두 분석 방법의 차이와 입력자료 특성에 기인한다. CWM 방법에서는 양의 값과 음의 값이 상 쇄되는 반면(식 7) GA 방법에서는 절대값이 사용되 므로 모두 더해진다(식 8). 이는 결과적으로 CWM에 서 기여도가 낮고 GA에서 기여도가 높은 경우 해당 기여도가 음의 가중치가 큰 것을 의미하며, 반대로 CWM에서 기여도가 높고 GA에서 기여도가 낮은 경 우 해당 기여도가 양의 가중치가 큰 것을 의미한다. 따라서 겨울철 서울의 Factor 3 기여도가 CWM 방법 에서 낮고 GA에서 높은 것은 음의 가중치의 역할에 의한 것임을 알 수 있다. 즉, Factor 3의 값이 작을 때 12시간 후 서울 PM2.5의 농도가 높아졌음을 지시하 며, 이는 종관기상 특성을 고려할 때 낮은 풍속으로

인한 대기의 정체가 12시간 후 서울 PM_{2.5} 농도에 높 은 기여를 보임을 의미한다. 일반적으로 겨울철 정체 시 온도와 습도는 높으며 풍향은 일정하지 않으므로 이들 인자 모두 음의 상관관계를 발생하기는 어렵다 (Lee *et al.*, 2007). 따라서 풍속이 Factor 3의 주요한 인 자로 작용하며, 결과적으로 낮은 풍속으로 인한 정체 효과가 겨울철 고농도 PM_{2.5} 발생에 큰 영향을 주는 것을 지시한다. 또한, 겨울철 울릉도 Factor 3의 기여 도가 CWM에서 가장 높았지만 GA에서 가장 낮은 것은 이러한 정체의 영향이 울릉도에서 가장 작은 것 을 의미한다.

Factor 3에서 CWM과 GA 방법의 차이는 종관기상 이 서울의 PM_{2.5} 질량농도를 결정하는 전반적인 조건 이지만 고농도 발생에는 정체의 영향이 중요함을 드 러낸다. 이는 현재의 예측 모델에서는 모사가 어려운 부분이므로, CWM과 GA와 같은 방법에 의한 신경망 모델의 입력 변수들의 기여도 평가는 서울과 같은 대 도시의 PM_{2.5} 특성을 이해하고 예측하는 데 활용 가 능한 유의미한 결과로 판단된다.

4. 결 론

본 연구에서는 2015~2019년 동안 여름(6~8월)과 겨울(12~2월)의 서울, 백령도, 제주도, 울릉도에서 생 산된 대기오염자료 및 기상자료를 활용하여 12시간 후의 서울시 초미세먼지(PM_{2.5}) 질량 농도를 예측하 는 인공신경망 모델을 구축하였다. 모델의 IOA는 여 름에는 0.81 겨울에는 0.86이었다. 그리고 이 모델을 바탕으로 가중치 연결방법(CWM)과 가르손 알고리 즘(GA) 방법을 통해 12시간 후 서울시 PM_{2.5} 질량 농 도에 대한 4 지역의 입력변수별 기여도를 정성적, 정 량적으로 산정하였다.

CWM과 GA분석 결과, 12시간 전 PM_{2.5} 농도인 Factor 1은 농도가 높은 겨울과 여름 모두 서울의 영 향이 가장 컸고, 백령도가 두 번째 순위였다. 이와는 반대로, O₃, NO₂, CO, SO₂ 등의 반응성 기체로 구성 된 Factor 2는 제주도와 울릉도의 기여도가 크게 나 타났다. 이러한 결과는 서울의 PM_{2.5}가 외부의 유입 에도 영향을 받으며, 서울과 그 주변에서 배출된 전 구기체에 의한 이차생성의 기여는 상대적으로 작음 을 의미한다. 기본 기상인자로 구성된 Factor 3는 CWM과 GA의 결과가 계절별로 차이를 보였는데 이 는 종관기상의 역할과 더불어 정체가 서울의 PM_{2.5} 를 결정하는 중요한 인자임을 지시한다.

CWM과 GA는 단일 은닉층에 대한 분석 방법이지 만, 이를 기반으로 본 연구에서 산출된 입력 변수의 기여도는 합리적인 결과로 판단된다. 단일 은닉층을 사용한 모델은 정확도가 비교적 높지 않아 이를 개선 하기 위한 여러 기법들(순환신경망의 사용 등)이 개 발되고 있지만 해석 방법은 추가적인 연구가 필요하 다. 추후 모델의 개발과 함께 본 연구와 같은 분석을 다양한 방법을 적용하여 수행한다면, 입력자료와 출 력자료 사이의 인과관계를 찾아 역동적으로 변하는 대기환경을 더 정확하게 이해하고 예측하는데 도움 이 될 것이다.

감사의 글

본 연구는 한국연구재단의 중견연구 (2020R1A2C 3014592) 과제의 지원을 받아 수행되었습니다.

References

- Adadi, A., Berrada, M. (2018) Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI), IEEE Access, 6, 52138-52160. https://doi.org/10.11 09/ACCESS.2018.2870052
- Bengio, Y., Grandvalet, Y. (2003) No unbiased estimator of the variance of K-fold cross-validation: Citeseer. https:// doi.org/10.1007/0-387-24555-3 5
- Chang, L.-S., Cho, A., Park, H., Nam, K., Kim, D., Hong, J.-H., Song, C.-K. (2016) Human-model hybrid Korean air quality forecasting system, Journal of the Air & Waste Management Association, 66(9), 896-911. https://doi.org/10.1080/10962247.2016.1206995
- Cho, K.-W., Jung, Y.-J., Kang, C.-G., Oh, C.-H. (2019a) Conformity

869

assessment of machine learning algorithm for particulate matter prediction, Journal of the Korea Institute of Information and Communication Engineering, 23(1), 20-26. https://doi.org/10.6109/ JKIICE.2019.23.1.20

- Cho, K., Lee, B.-Y., Kwon, M., Kim, S. (2019) Air Quality Prediction Using a Deep Neural Network Model, Journal of Korean Society for Atmospheric Environment, 35(2), 214-225. https://doi.org/10.5572/KOSAE. 2019.35.2.214
- Garson, G.D. (1991) A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data, Social Science Computer Review, 9(3), 399-434. https://doi.org/10.1177/089443939100900304
- Goh, A.T. (1995) Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering, 9(3), 143-151. https://doi.org/10.1016/0954-1810(94)00011-S
- Ho, C.H., Park, I., Oh, H.-R., Gim, H.-J., Hur, S.-K., Kim, J., Choi, D.-R. (2021) Development of a PM_{2.5} prediction model using a recurrent neural network algorithm for the Seoul metropolitan area, Republic of Korea, Atmospheric Environment, 245, 118021. https://doi. org/10.1016/j.atmosenv.2020.118021
- Hwang, K.-W., Kim, D.-Y., Jin, S.-J., Kim, I.-H. (2020) A study on the factors influencing air pollutions in the islands of Korean penisula: Focusing on the case of Ulleung, Jeju, and Baengnyong Island, Journal of the Korea Academia-Industrial cooperation Society, 21(11), 814-824. https://doi.org/10.5762/KAIS.2020. 21.11.814
- Kim, J.-A., Lim, S., Shang, X., Lee, M., Kang, K.-S., Ghim, Y.S. (2020) Characteristics of PM_{2.5} chemical composition and high-concentration Episodes observed in Jeju from 2013 to 2016, Journal of Korean Society for Atmospheric Environment, 36(3), 388-403. https://doi.org/10.5572/KOSAE.2020.36.3.388
- Kim, Y.P. (2006) Air pollution in Seoul caused by aerosols, Journal of Korean Society for Atmospheric Environment, 22(5), 535-553.
- Lee, J.-Y., Han, J.-S., Kong, B.-J., Hong, Y.-D., Lee, J.-H., Chung, I.-R. (2007) Variation of PM₁₀ concentration in Seoul in association with synoptic meteorological conditions, Journal of Environmental Impact Assessment, 16(5), 351-361.
- Lim, J.-H., Ahn, J.-Y., Seo, S.-J., Seo, Y.-K., Hong, Y.-D., Han, J.-S. (2016) Source Identification and Apportionment of

PM_{2.5} in Baengnyeong Island, Korea in 2015, Journal of the Korean Society of Urban Environment, 16(4), 461-471.

- Lim, J.-H., Park, J.-S., Ahn, J.-Y., Choi, J.-S., Oh, J., Moon, K.-J., Hong, Y.-D., Han, J.-S. (2013) The Characteristics of the Air Pollutants at Baengnyeong Island, a West Inflow Region of the Korean Peninsula, Journal of the Korean Society of Urban Environment, 13(3), 267-276.
- Moon, K.J., Han, J.S., Kong, B.J., Lee, M.D., Jung, I.R. (2005) Characteristics of chemical species in gaseous and aerosol phase measured at Gosan, Korea during ABC-EAREX2005, Journal of Korean Society for Atmospheric Environment, 21(6), 675-687.
- Olden, J.D., Jackson, D.A. (2002) Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks, Ecological Modelling, 154(1-2), 135-150. https://doi. org/10.1016/S0304-3800(02)00064-9
- Olden, J.D., Joy, M.K., Death, R.G. (2004) An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, Ecological Modelling, 178(3-4), 389-397. https://doi. org/10.1016/j.ecolmodel.2004.03.013
- Park, K.Y., Lee, H.G., Suh, M.S., Jang, K.M., Kang, C.-H., Hu, C.-G., Kim, Y.-J. (1994) Analysis of Air Pollution Concentrations at Cheju Baseline Measurement Station, Journal of Korean Society for Atmospheric Environment, 10(4), 252-259.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. (2019) Deep learning and process understanding for data-driven Earth system science, Nature, 566(7743), 195-204. https://doi.org/10.1038/ s41586-019-0912-1
- Wang, Y.L., Song, W., Yang, W., Sun, X.C., Tong, Y.D., Wang, X.M., Liu, C., Bai, Z., Liu, X.Y. (2019) Influences of atmospheric pollution on the contributions of major oxidation pathways to PM_{2.5} nitrate formation in Beijing, Journal of Geophysical Research: Atmospheres, 124(7), 4174-4185. https://doi.org/10.1029/2019 JD030284

Authors Information

길준수(고려대학교 지구환경과학과 석박통합과정) 이미혜(고려대학교 지구환경과학과 교수)